European Signal Processing Conference (EUSIPCO)
Vienna, Austria, September 2004

EXTENSION AND EVALUATION OF MDS FOR GEOMETRIC MICROPHONE ARRAY CALIBRATION

Amarnag Subramanya and Stanley T. Birchfield

Dept. of Electrical and Computer Engineering
Clemson University, Clemson, SC 29634
{asubram,stb}@clemson.edu

ABSTRACT

Classical multidimensional scaling (MDS) is a simple, global, non-iterative technique for determining the locations of microphones given their interpoint distances, even when all such distances are not available. We extend this technique by showing how it can be used with a calibration target consisting of synchronized sound sources, which facilitates automatic calibration of large arrays when such a target is available. We also evaluate the sensitivity of the algorithm to the size and location of the target, as well as to the basis points in general, on several different microphone array configurations. Simulations and experiments demonstrate the accuracy of the algorithm for practical scenarios, yielding errors on the order of 1 cm.

1. INTRODUCTION

Applications using microphone arrays, such as acoustic localization and beamforming, require the locations of the microphones to be known. Determining these locations is the problem of geometric microphone array calibration. Traditional methods of calibrating microphone arrays have involved expensive calibration targets and/or nonlinear optimization techniques that are subject to local minima [7, 1].

Recently it was shown [2] that the locations of the microphones can be computed using a simple, global, non-iterative technique known as *classical multidimensional scaling (MDS)* [5]. Classical MDS is an algorithm which, given noisy distances between a set of points in a Euclidean space, estimates the coordinates of those points. The algorithm is robust to measurement noise, yielding accuracy on the order of the smallest audible wavelength[2]. If measuring all the inter-microphone distances is impractical (e.g., with a large microphone arrays [8]), an alternative is to apply classical MDS to the distances between each microphone and a set of basis points [2].

In this paper we extend the algorithm to use a calibration target consisting of four sound sources. We measure the sensitivity of the algorithm to the size and position of the target on several different microphone array configurations. In addition, we measure its sensitivity to the choice of basis points in general. Simulations and experiments verify the accuracy of the algorithm for practical applications, with errors on the order of 1 cm.

To compute the locations of n microphones in a p-dimensional space from their pairwise distances δ_{ij},

1. Construct the squared-distance matrix D whose entries are δ_{ij}^2.
2. Compute the inner product matrix $B = -\frac{1}{2}JDJ$, where $J = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is the double-centering matrix and $\mathbf{1}$ is a vector of all ones.
3. Decompose B as $B = V\Lambda V^T$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, the diagonal matrix of eigenvalues of B, and $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, the matrix of corresponding unit eigenvectors. Sort the eigenvalues in non-increasing order: $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$.
4. Extract the first p eigenvalues $\Lambda_p = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ and corresponding eigenvectors $V_p = [\mathbf{v}_1, \ldots, \mathbf{v}_p]$.
5. The microphone coordinates are now located in the $n \times p$ matrix $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T = V_p \Lambda_p^{\frac{1}{2}}$.

Figure 1: Classical multidimensional scaling algorithm.

2. CLASSICAL MULTIDIMENSIONAL SCALING

Suppose we have n microphones in a p-dimensional space (usually $p = 3$) and have measured the distance $\delta_{ij} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)}$ between each pair of microphones i and j. The classical multidimensional scaling algorithm, given in Figure 1, computes the optimal locations of the microphones, in the sense that $\sum_{i=1}^n \sum_{j=1}^n (\delta_{ij} - \hat{\delta}_{ij})^2$ is minimized.

If measuring all the pairwise distances is prohibitive, an alternative is to use a set of $p+1$ basis points [2]. These basis points may be a subset of the microphone locations or any other points in the space (e.g., the sound source locations on a calibration target). The algorithm works as follows. First, the coordinates of the basis points are computed from their pairwise distances using the equations in Figure 2. Then, those equations are used to compute *preliminary coordinates* for each microphone given the distances between the basis points and the microphone. These preliminary coordinates are used to fill the squared distance matrix D which is then fed to the classical MDS algorithm.

For brevity we will refer to the classical MDS algorithm (Figure 1) as $\mathcal{A}_o$and the version using basis points (Figures 1 and 2) as $\mathcal{A}_b$.

In three dimensions, the coordinates of the four basis points are given by

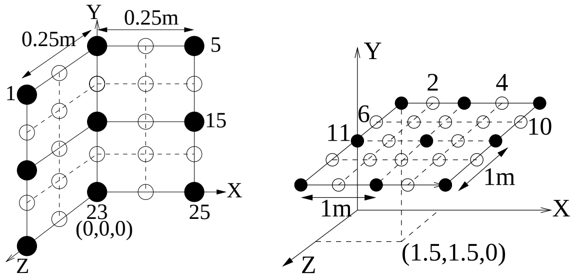
$$
\begin{aligned}
A &: \left(-\frac{1}{2}\delta_{AB},\ 0,\ 0\right) \\
B &: \left(\frac{1}{2}\delta_{AB},\ 0,\ 0\right) \\
C &: \left(x_C,\ \sqrt{\frac{1}{2}\delta^2_{AC} - \frac{1}{4}\delta^2_{AB} + \frac{1}{2}\delta^2_{BC} - x_C^2},\ 0\right) \\
D &: \left(x_D,\ y_D,\ \sqrt{\frac{1}{2}\delta^2_{AD} - \frac{1}{4}\delta^2_{AB} + \frac{1}{2}\delta^2_{BD} - x_D^2 - y_D^2}\right).
\end{aligned}
$$

The coordinates of an arbitrary point Q can be obtained by its distance to the basis points:

$$
\begin{aligned}
x_Q &= \frac{\delta^2_{AQ} - \delta^2_{BQ}}{2\delta_{AB}} \\
y_Q &= \frac{\delta^2_{AC} - \delta^2_{AB} + \delta^2_{BC} + \delta^2_{AQ} + \delta^2_{BQ} - 2\delta^2_{CQ} - 4x_C x_Q}{4y_C} \\
z_Q &= \frac{1}{4z_D}\left(\delta^2_{AD} + \delta^2_{BD} + \delta^2_{AQ} + \delta^2_{BQ} - \delta^2_{AB}\right. \\
&\qquad \left. -2\delta^2_{DQ} - 4x_D x_Q - 4y_D y_Q\right).
\end{aligned}
$$

Note that x_C, x_D, and y_D are found by substituting C or D for Q.

Figure 2: Equations for computing and using basis point coordinates.



(a) Ω_1 (all 25 mics)
Ω_1' (9 dark mics)

(b) Ω_2 (all 25 mics)
Ω_2' (9 dark mics)

Figure 3: The two microphone array configurations.

3. SENSITIVITY ANALYSIS

In this section we evaluate the sensitivity of $\mathcal{A}_b$ to the choice of basis points.

3.1 Computation of RMS Error

In these simulations, the ground truth microphone locations were stored in an $n \times p$ matrix Y, while the result of the algorithm was stored in an $n \times p$ matrix X. Since the solution has an arbitrary translation, rotation, and reflection, we measured the root-mean-square (RMS) error in the solution as $\sqrt{R^2}/n$, where

$$R^2 = \sum_{i=1}^{n} (A^T \mathbf{x_i} - \mathbf{y_i})^T (A^T \mathbf{x_i} - \mathbf{y_i}),$$

$\mathbf{y_i^T}$ is the ith row of Y, $\mathbf{x_i^T}$ is the ith row of X, and A is the normalized orthogonal matrix that best aligns the points by rotating and reflecting them [2, 5]:

$$A = (X^T Y Y^T X)^{\frac{1}{2}} (Y^T X)^{-1}.$$

Translation was handled by first shifting all the points so that their centroid coincided with the origin.

3.2 Array Configurations

For our analysis, we considered two microphone array configurations Ω_1 and Ω_2 similar to those used by Brandstein et al. [3] and Checka et al. [4], respectively, shown in Figure 3. In addition to the nine microphones contained in the arrays of those papers (shown as dark circles in the figure), we inserted the additional microphones shown as hollow circles, bringing the total number of microphones per configuration to 25. The original 9-microphone configurations are denoted by Ω_1' and Ω_2'.

With the 9 microphones of Ω_1' or Ω_2', $\mathcal{A}_o$ requires $n(n-1)/2 = 36$ pairwise distances while $\mathcal{A}_b$ requires $p(p+1)/2 + (p+1)(n-p-1) = 26$. In such a case, $\mathcal{A}_b$ does not provide much savings in terms of measurement labor, and $\mathcal{A}_o$ is probably the method of choice. In contrast, with the 25 microphones of Ω_1 or Ω_2, $\mathcal{A}_b$ requires just 90 distances compared with 300 distances for $\mathcal{A}_o$ — a savings of 70%.

3.3 Choice of Basis Points

As explained in [2], the robustness of the algorithm $\mathcal{A}_b$ increases with the volume enclosed by the basis points (or the area in the case of a planar array). To use $\mathcal{A}_b$ with a subset of the inter-microphone distances, then, one should select as basis points the microphones whose enclosed volume (or area) is maximized.

To test the sensitivity of the algorithm to the choice of basis points, we measured the accuracy of $\mathcal{A}_b$ using all possible $\binom{n}{p+1}$ basis sets.[1] We perturbed the distance measurements by adding a zero-mean additive Gaussian noise with $\sigma = 1$ cm. Then we ran $\mathcal{A}_b$ on those measurements, averaging over 100 trials, obtaining the results shown in Table 1. For both configurations the best choice of basis points yielded an accuracy much smaller than the smallest audible wavelength (approximately 1.6 cm). Because of the additional constraint afforded by the reduced dimensionality of the planar array Ω_2, however, its accuracy is greater than that of the three-dimensional array Ω_1.

Manually selecting good basis points for common microphone array configurations is not difficult. Looking at Figure 3, the microphones $\langle 1, 5, 23, 25\rangle$ (or other symmetric choices) clearly maximize the volume of Ω_1. Similarly, with Ω_2 the microphones $\langle 1, 5, 23\rangle$ maximize the area. The error for these two choices are 0.3 cm and 0.2 cm, respectively, which are statistically the same as the best results.

[1] But we excluded sets whose basis points A, B, and C were collinear or A, B, C, and D were coplanar (for Ω_1), which would cause a divide-by-zero error in the equations of Figure 2.

configuration	mic indices	RMS error (cm)
Ω_1	$\langle 1,5,22,24\rangle$	0.3
	$\langle 1,3,5,19\rangle$	0.3
	$\langle 13,17,22,24\rangle$	5.6
	$\langle 12,14,17,19\rangle$	5.9
Ω_2	$\langle 4,16,24\rangle$	0.2
	$\langle 2,20,21\rangle$	0.2
	$\langle 3,7,9\rangle$	3.9
	$\langle 5,13,14\rangle$	4.0

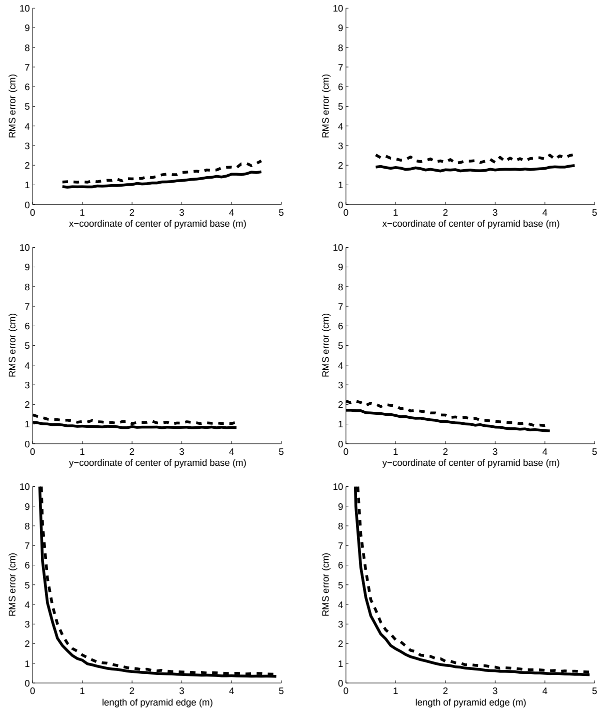
Table 1: The two best and two worst unique basis sets for Ω_1 and Ω_2. Note that other basis sets have similar errors because of symmetry.

4. USING A CALIBRATION TARGET

Up to now we have assumed that the basis points are a subset of the microphones and that the interpoint distances are measured with a tape measure. Since $\mathcal{A}_b$ works for any basis points in the space, however, an alternative is to use a calibration target consisting of four speakers rigidly attached to one another, with the four speakers being the basis points. By measuring the $p(p+1)/2 = 6$ inter-speaker distances with a tape measure and the $n(p+1)$ distances between microphones and speakers using time-delay estimation, $\mathcal{A}_b$ yields a simple, non-iterative technique that globally computes the microphone locations. Other well-known techniques, such as the simplex method or stochastic region contraction (SRC) [7, 1], solve a nonlinear equation by iterating around an initial solution and thus are subject to local minima.

Measuring the distances between speakers and microphones requires that we have access to the reference signal emitted by each speaker, synchronized with the signals received by the microphones. One way to achieve this prerequisite is to synchronize the speaker emission with the microphone digitizers (e.g., by modulating a radio-frequency (RF) signal that travels instantaneously compared with the sound wave [6]) and to use the waveform sent to the speaker as the reference. Another way is to place an extra microphone as close as possible to each speaker to capture the sound as soon as it is emitted. This extra microphone, whose digitizer must be synchronized with the other microphones' digitizers, provides the reference signal. A third alternative is to synchronize the signals . Either way, time-delay estimation cross-correlates the reference signal with the signals received by the microphones to estimate the time it took sound to travel from the source to the microphone. For a good description of the difficulties and details involved in using a calibration target, including the complications arising from changes in the speed of sound, please consult Sachar et al. [7].

For our purposes we assume that these problems are solved, and that the distances between speakers and microphones can be measured. In this subsection we wish to evaluate the sensitivity of $\mathcal{A}_b$ to the location and size of the calibration target and to gain insight regarding the design choice for these parameters. In order to maximize the volume enclosed by the basis points, the shape of the calibration target was chosen to be a right, regular, triangle-based pyramid.



(a) Ω_1 (solid) and Ω_1' (dashed)

(b) Ω_2 (solid) and Ω_2' (dashed)

Figure 4: RMS error for the two configurations versus y-coordinate (TOP), x-coordinate (MIDDLE), and size of calibration target (BOTTOM).

For the configurations Ω_1 and Ω_1', we placed the pyramidal target in the center of a $5 \times 5 \times 5$ m room, resting on the floor, so that the coordinates of the center of its base were $(2.5, 0, 2.5)$ m. First we uniformly scaled the target, varying the pyramidal edge length from 0.1 m to 5 m in increments of 0.1 m. Then, setting the length to 1 m, we varied the target's x-coordinate, sliding the target from one end of the room to the other. Finally, setting the x-coordinate to 2.5 m, we varied the y-coordinate in a similar manner. This entire procedure was then repeated for the configurations Ω_2 and Ω_2'. In all simulations we required the calibration target to remain inside the room, and all distance measurements were corrupted by zero-mean Gaussian noise with $\sigma = 1$ cm.

Several conclusions can be drawn from these results, which are displayed in Figure 4. As expected, the error decreased with the size of the calibration target, reaching the order of the smallest audible wavelength (1.6 cm) when the pyramidal edge was about 1 m. Position had less effect on the results, but in all cases the error decreased as the target moved closer to the microphone array. Not surprisingly, the error of the three-dimensional array Ω_1 was less than that of the planar array Ω_2 because the computation in both cases was performed in 3D. Most importantly, with a sufficiently-sized target the errors ranged from approximately 1 to 2 cm (standard deviation between 0.4 and 0.6 cm when pyramidal edge length is at least 1 m), which validates the accuracy of this technique for practical acoustic applications.

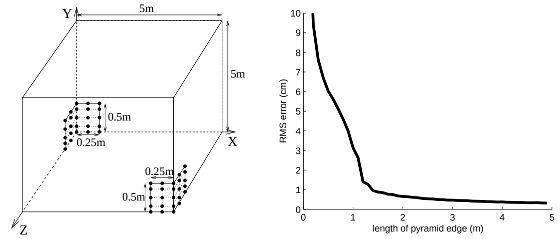


Figure 5: Two Ω_1s placed at opposite corners of a room, along with the error vs. the size of the calibration target.

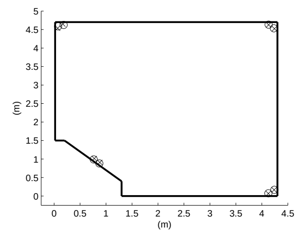


Figure 6: The results of a real experiment. Microphone locations computed by the algorithm $\mathcal{A}_o$ (`x`), and those computed by measuring by hand (`o`).

5. DETACHED MICROPHONES

Although the simulations of the previous section validate the accuracy of the algorithm and provide insight as to the effect that various parameters have on the outcome, the arrays Ω_1 and Ω_2 themselves are simplistic examples. In practice the microphone coordinates of such rectilinear arrays would be known at the time the array was constructed. Nevertheless, the results of these simulations are applicable to similarly-sized and similarly-shaped arrays which are not constructed so carefully, for example, a distributed set of laptop computers each containing a built-in microphone, as in [6].

More generally, when microphones are not rigidly connected to one another, a calibration technique is necessary because their locations are not known *a priori*. For example, even though the microphone locations may be known within a rigidly-attached array, when two or more such arrays are placed in a room the locations of all the microphones are not known with respect to a single coordinate system. In Figure 5 we have placed two arrays of Ω_1 at opposite corners of the room and varied the size and location of the calibration target as before, using $\mathcal{A}_b$ to compute the microphone locations.[2] The results are similar, though the drop-off due to size is sharper because of the increased proximity of the target to both arrays as the size increased. As expected, the error was minimum with the target in the middle of the room, although the figures for x- and y-coordinates were omitted from this paper due to space constraints.

This work was originally motivated by the real-world scenario shown in Figure 6. Eight microphones were located in a semi-rectangular conference room, a pair of microphones in each corner of the room. Although the intra-pair distance was known *a priori* to be 15 cm, the distances between pairs of microphones were unknown. Measuring the 24 inter-microphone distances (in addition to the *a priori* distances) and using $\mathcal{A}_o$ yielded the results shown in the figure. The RMS difference between the results and the manual measurements was 2.6 cm.

6. CONCLUSION

We have shown how to apply the MDS algorithm to a calibration target with four sound sources, thus facilitating automatic calibration when such a target is available. A large target ($\geq$ 1 m edge length) is necessary to achieve good results, but the algorithm is fairly insensitive to the target's position. Choosing basis points in general on a microphone array is not difficult, with excellent results possible by manually selecting points that maximize the enclosed volume (or area, in the case of a planar array). Simulations and experiments have demonstrated accuracy on the order of 1 cm, which verifies the algorithm's applicability to practical scenarios.

[2] Because real rooms are not constructed with perfectly straight lines, the problem in reality is harder than it appears in the figure.

REFERENCES

[1] M. Berger and H. F. Silverman. Microphone array optimization by stochastic region contraction (SRC). *IEEE Transactions on Signal Processing*, 39(11):2377–2386, Nov. 1991.

[2] S. T. Birchfield. Geometric microphone array calibration by multidimensional scaling. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003.

[3] M. S. Brandstein, J. E. Adcock, and H. F. Silverman. A closed-form method for finding source locations from microphone-array time-delay estimates. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 3019–3022, 1995.

[4] N. Checka, K. Wilson, V. Rangarajan, and T. Darrell. A probabilistic framework for multi-modal multi-person tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[5] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. London: Chapman and Hall, 1994.

[6] V. Raykar, I. Kozintsev, and R. Lienhart. Position calibration of audio sensors and actuators in a distributed computing platform. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, pages 572–581, Nov. 2003.

[7] J. M. Sachar, H. F. Silverman, and W. R. P. III. Position calibration of large-aperture microphone arrays. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002.

[8] H. F. Silverman, W. R. Patterson, J. L. Flanagan, and D. Rabinkin. A digital processing system for source location and sound capture by large microphone arrays. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997.